[0051] The weight of a term is not necessarily a positive value. If a term has a negative weight, it may suggest that the user prefers that his search results should not include this term and the magnitude of the negative weight indicates the strength of the user's preference for avoiding this term in the search results. By way of example, for a user who is breeds Australian Shepard dogs in San Francisco, Calif., the term-based profile may include terms like "Australian Shepard", "agility training" and "San Francisco" with positive weights. The terms like "German Shepard" or "Australia" may also be included in the profile. However, these terms are more likely to receive a negative weight since they are irrelevant and confusing with the authentic preference of this particular user.

[0052] A term-based profile itemizes a user's preference using specific terms, each term having certain weight. If a document contains a term that is in a user's term-based profile, the term's weight will be assigned to the document; however, if a document does not contain the term, it will not receive any weight associated with this term. Such a requirement of relevance between a document and a user profile sometimes may be less flexible when dealing with various scenarios in which a fuzzy relevance between a user's preference and a document exists. For example, if a user's term-based profile includes terms like "Mozilla" and "browser", a document containing no such terms, but other terms like "Galeon" or "Opera" will not receive any weight because they do not match any existing term in the profile, even though they are actually Internet browsers. To address the need for matching a user's interests without exact term matching, a user's profile may include a category-based profile.

[0053] **FIG. 4A** illustrates a hierarchical category map **400** according to the Open Directory Project (http://dmoz.org/). Starting from the root level of map **400**, documents are organized under several major topics, such as "Art", "News", "Sports", etc. These major topics are often too broad to delineate a user's specific interest. Therefore, they are further divided into sub-topics that are more specific. For example, topic "Art" may comprise sub-topics like "Movie", "Music" and "Literature" and the sub-topic "Music" may further comprise sub-sub-topics like "Lyrics", "News" and "Reviews". Note that each topic is associated with a unique CATEGORY_ID like 1.1 for "Art", 1.4.2.3 for "Talk Show" and 1.6.1 for "Basketball".

[0054] A user's specific interests may be associated with multiple categories at various levels, each of which may have a weight indicating the degree of relevance between the category and the user's interest. In one embodiment, a category-based profile may be implemented using a hash table data structure as shown in **FIG. 4B**. A category-based profile table **450** includes a table **455** that comprises a plurality of records **460**, each record including a USER_ID and a pointer pointing to another data structure, such as table **460-1**. Table **460-1** may include two columns, CATEGORY_ID column **470** and WEIGHT column **480**. CATEGORY_ID column **470** contains a category's identification number as shown in **FIG. 4A**, suggesting that this category is relevant to the user's interests and the value in the WEIGHT column **480** indicates the degree of relevance of the category to the user's interests.

[0055] A user profile based upon the category map **400** is a topic-oriented implementation. The items in a category-based profile can also be organized in other ways. In one embodiment, a user's preference can be categorized based on the formats of the documents identified by the user, such as HTML, plain text, PDF, Microsoft Word, etc. Different formats may have different weights. In another embodiment, a user's preference can be categorized according to the types of the identified documents, e.g., an organization's homepage, a person's homepage, a research paper, or a news group posting, each type having an associated weight. Another type category that can be used to characterize a user's search preferences is document origin, for instance the country associated with each document's host. These types of category information can be derived from either the user's prior searches **203**, or from the user's web related information **217**. In yet another embodiment, the above-identified category-based profiles may co-exist, with each one reflecting one aspect of a user's preferences.

[0056] Besides term-based and category-based profiles, another type of user profile is referred to as a link-based profile. As discussed above, a page rank algorithm, such as disclosed in U.S. Pat. No. 6,285,999 uses the link structure that connects various documents over the Internet. A document that has more links pointing to it is often assigned a higher page rank and therefore attracts more attention from a search engine. Link information related to a document identified by a user can also be used to infer the user's preferences. In one embodiment, a list of preferred URLs are identified for a user by analyzing the frequency of his access to those URLs. Each preferred URL may be further weighted according to the time spent by the user and the user's scrolling activity at the URL, and/or other user activities **209** when visiting the document at the URL. In another embodiment, a list of preferred hosts are identified for a user by analyzing the user's frequency of accessing web pages of different hosts. When two preferred URLs are related to the same host the weights of the two URLs may be combined to determine a weight for the host. In another embodiment, a list of preferred domains are identified for a user by analyzing the user's frequency of accessing web pages of different domains. For example, for finance.yahoo.com, the host is "finance.yahoo.com" while the domain is "yahoo.com".

[0057] **FIG. 5** illustrates a link-based profile using a hash table data structure. A link-based profile table **500** includes a table **510** that includes a plurality of records **520**, each record including a USER_ID and a pointer pointing to another data structure, such as table **510-1**. Table **510-1** may include two columns, LINK_ID column **530** and WEIGHT column **540**. The identification number stored in the LINK_ID column **530** may be associated with a preferred URL or host. The actual URL/host/domain may be stored in the table instead of the LINK_ID, however it is preferable to store the LINK_ID to save storage space.

[0058] A preferred list of URLs and/or hosts includes URLs and/or hosts that have been directly identified by the user. The preferred list of URLs and/or host may furthermore extend to URLs and/or hosts indirectly identified by using methods such as collaborative filtering or bibliometric analysis, which are known to persons of ordinary skill in the art. In one embodiment, the indirectly identified URLs and/or host include URLs or hosts that have links to/from the directly identified URLs and/or hosts. These indirectly identified URLs and/or hosts are weighted by the distance